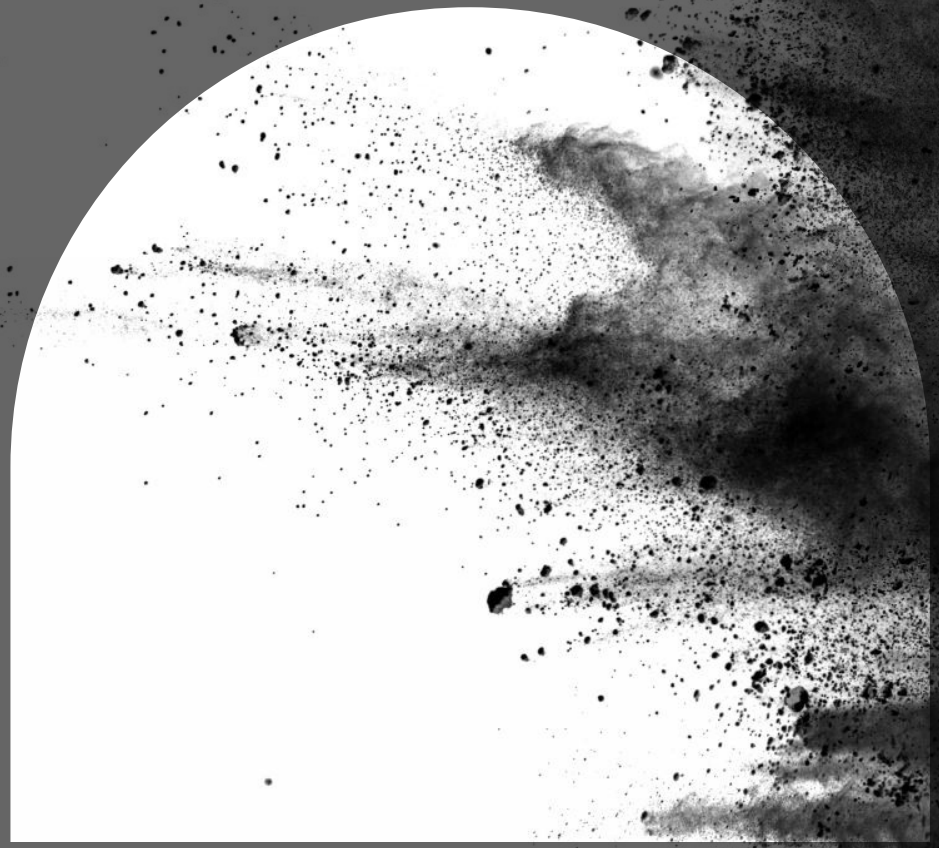


PREDICTING MOVIE REVENUE USING REGRESSION

- Timothy Tsung

Objective

- To predict movie gross revenue based on past movies and corresponding data points collected from box office mojo

Why?

- New film company DDT wants to know what type of movie they should film next?

Target vs. Features

DOMESTIC GROSS REVENUE

MOVIE RUN TIME, WIDEST THEATRE RELEASE, GENRE, MPAA RATING

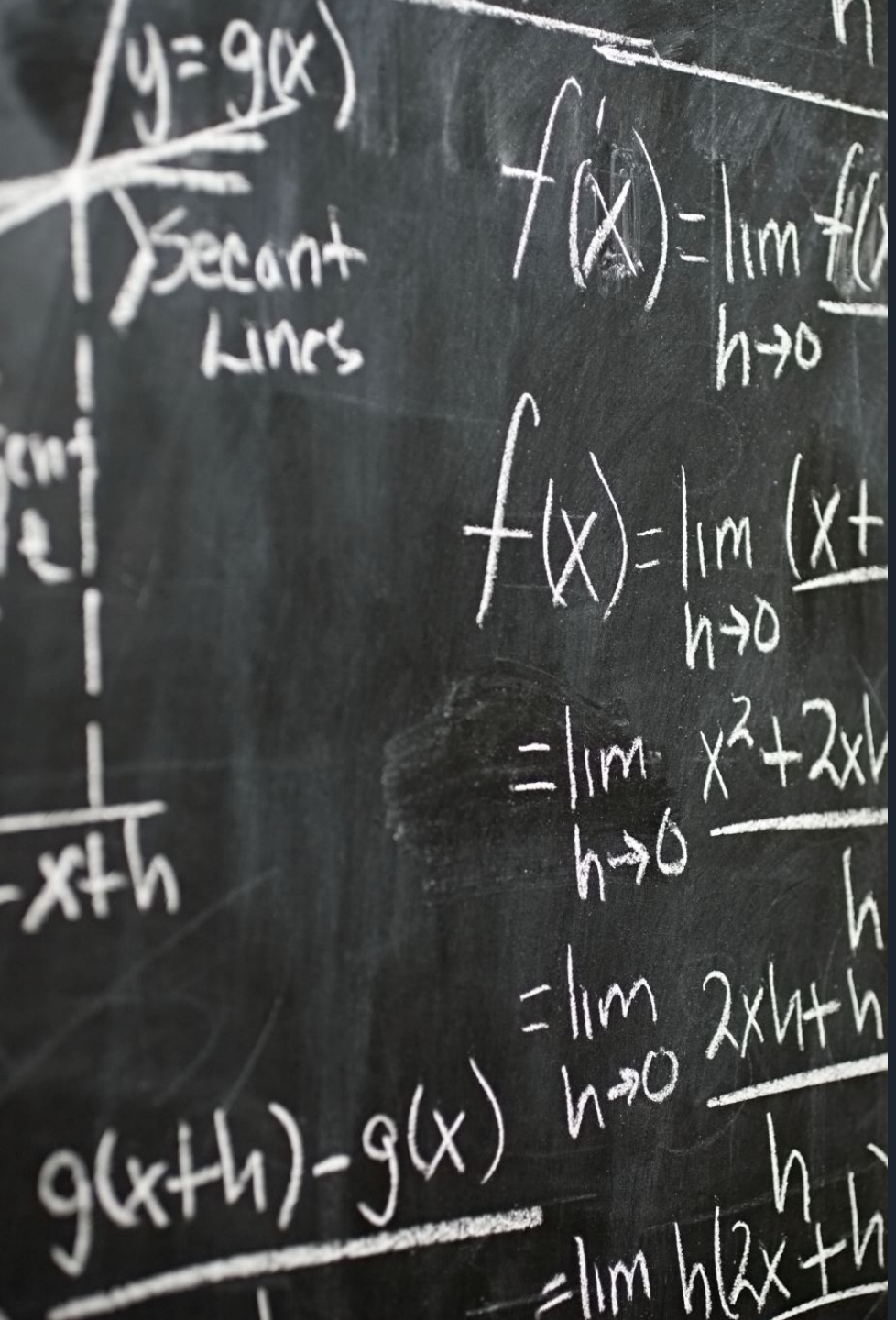
Methods

- Feature engineering to convert categorical data into binominal dummy sets
- Test for best fit model:

  Linear regression

  Ridge regression

  Polynomial regression degree 2



Results

Train/test split (R^2)

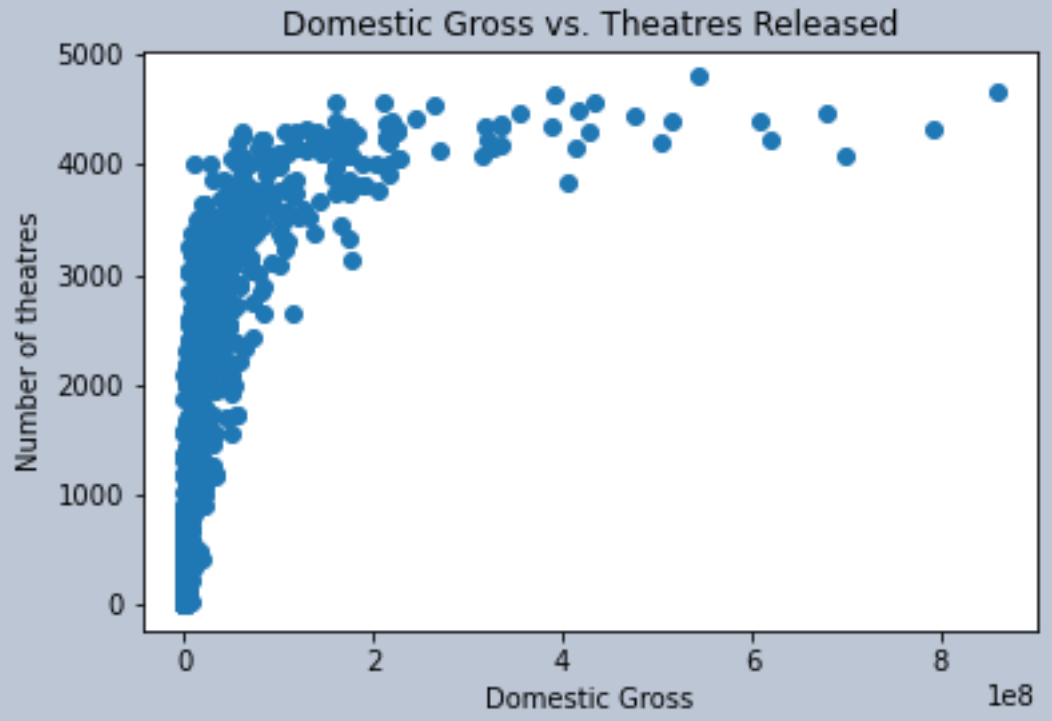
- Linear Regression = 0.302
- Ridge Regression = 0.302
- Polynomial Regression = 0.507

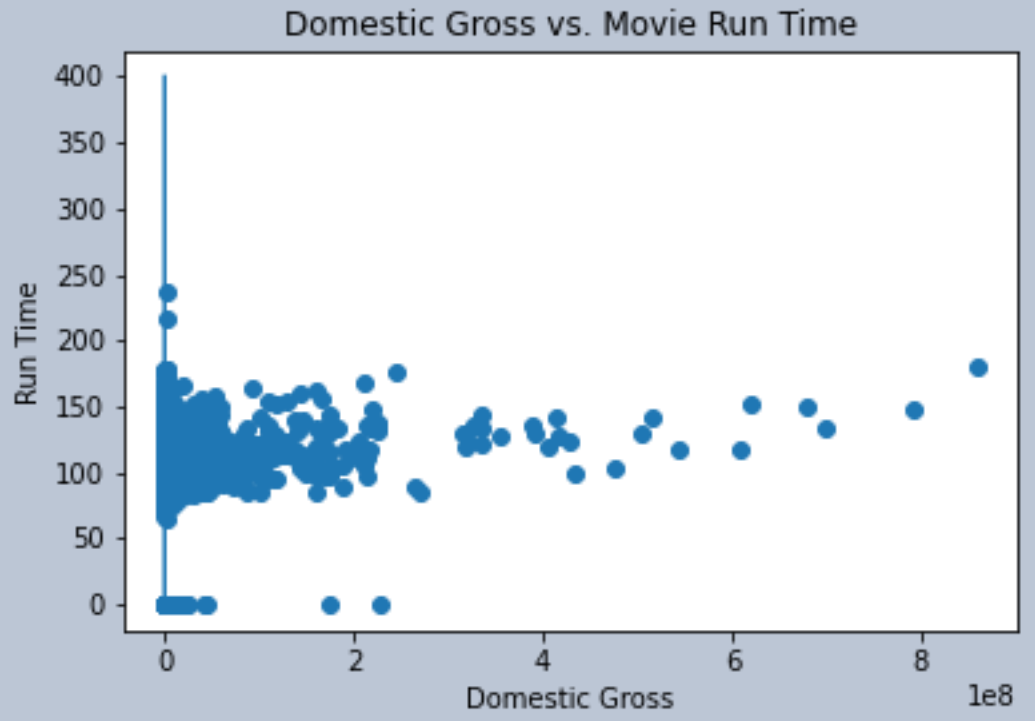
K-Fold Cross Validation (R^2)

- Linear Regression = 0.3808
- Ridge Regression = 0.3814
- Polynomial Regression = 0.3808

Results

- Correlation indicated [widest release theatres] with highest relationship (0.61)
- Interested to see the relationship with movie run time





Conclusion

- Best fit model is ridge regression
- Final score → 0.302
- Very low R^2, and it means roughly 30% of the domestic gross revenue's variance can be explained by the features I selected

Therefore

- Not confident my model can predict an accurate gross revenue for DDT

Future

- Add more engineered features than just genre and MPAA rating
- Explore other features like release date/seasons
- Play with other types of models & train/validate/test combinations

Questions?